SSOAR Open Access Repository
www.ssoar.info

On the feasibility of portfolio optimization under expected shortfall

Ciliberti, Stefano; Kondor, Imre; Mézard, Marc

Postprint / Postprint
Zeitschriftenartikel / journal article

Zur Verfügung gestellt in Kooperation mit / provided in cooperation with:
www.peerproject.eu

Empfohlene Zitierung / Suggested Citation:
Ciliberti, S., Kondor, I., & Mézard, M. (2007). On the feasibility of portfolio optimization under expected shortfall. *Quantitative Finance*, *7*(4), 389-396. https://doi.org/10.1080/14697680701422089

Nutzungsbedingungen:
Dieser Text wird unter dem "PEER Licence Agreement zur Verfügung" gestellt. Nähere Auskünfte zum PEER-Projekt finden Sie hier: http://www.peerproject.eu Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.
Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use:
This document is made available under the "PEER Licence Agreement ". For more Information regarding the PEER-project see: http://www.peerproject.eu This document is solely intended for your personal, non-commercial use.All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.
By using this particular document, you accept the above-stated conditions of use.

gesis Leibniz-Institut für Sozialwissenschaften

Mitglied der Leibniz-Gemeinschaft

Diese Version ist zitierbar unter / This version is citable under:
https://nbn-resolving.org/urn:nbn:de:0168-ssoar-221061

QUANTITATIVE FINANCE

On the Feasibility of Portfolio Optimization under Expected Shortfall

Journal:	*Quantitative Finance*
Manuscript ID:	RQUF-2006-0108.R1
Manuscript Category:	Research Paper
Date Submitted by the Author:	28-Jan-2007
Complete List of Authors:	ciliberti, stefano; univ paris sud, LPTMS Kondor, Imre Mezard, Marc
Keywords:	Portfolio Optimization, Risk Measures, Statistical Physics, Financial Applications
JEL Code:	
Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.	
resubmission.tgz	

SCHOLARONE™
Manuscripts

For Peer Review Only

On the Feasibility of Portfolio Optimization under Expected Shortfall

Stefano Ciliberti,[1] Imre Kondor,[2] and Marc Mézard[1]
[1]CNRS; Univ. Paris Sud, UMR8626, LPTMS, ORSAY CEDEX, F-91405
[2]Collegium Budapest, 1014 Budapest, Szenthromsg u. 2

We address the problem of portfolio optimization under the simplest coherent risk measure, i.e. the expected shortfall. As it is well known, one can map this problem into a linear programming setting. For some values of the external parameters, when the available time series is too short, the portfolio optimization is ill posed because it leads to unbounded positions, infinitely short on some assets and infinitely long on some others. As first observed by Kondor and coworkers, this phenomenon is actually a phase transition. We investigate the nature of this transition by means of a replica approach.

I. INTRODUCTION

Among the several risk measures existing in the context of portfolio optimization, expected shortfall (ES) has certainly gained increasing popularity in recent years. In several practical applications, ES is starting to replace the classical Value-at-Risk (VaR). There are a number of reasons for this. For a given threshold probability β, the VaR is defined so that with probability β the loss will be smaller than VaR. This definition only gives the minimum loss one can reasonably expect but does not tell anything about the typical value of that loss that can be measured by the *conditional* value-at-risk (CVaR, which is the same as ES for continuous distributions that we consider here [12]). We will be more precise on these definitions below. The point we want to stress here is that the VaR measure, lacking the mandatory properties of subadditivity and convexity,is not *coherent* [1]. This means that summing VaR's of individual portfolios will not necessarily produce an upper bound for the VaR of the combined portfolio, thus contradicting the holy principle of diversification in finance. A nice practical example of the inconsistency of VaR in credit portfolio management is reported in Ref. 3. On the other hand, it has been shown [2] that ES is a coherent measure with interesting properties [4]. Moreover, the optimization of ES can be reduced to linear programming [5] (which allows for a fast implementation) and leads to a good estimate for the VaR as a byproduct of the minimization process. To summarize, the intuitive and simple character, together with the mathematical properties (coherence) and the fast algorithmic implementation (linear programming), are the main reasons behind the growing importance of ES as a risk measure.

In this paper, we will focus on the feasibility of the portfolio optimization problem under the ES measure of risk. The control parameters of this problem are (i) the imposed threshold in probability, β, and (ii) the ratio N/T between the number N of financial assets making up the portfolio and the time series length T used to sample the probability distribution of returns. (It is curious that, albeit trivial, the scaling in N/T had not been explicitly pointed out before [11]. It has been discovered by Kondor at al. Ref. 6 that, for certain values of these parameters, the optimization problem does not have a finite solution because, even if convex, it is not bounded from below. Extended numerical simulations allowed these authors to determine the feasibility map of the problem. Here, in order to better understand the root of the problem and to study the transition from a feasible regime to an unfeasible one (corresponding to an ill-posed minimization problem) we address the same problem from an analytical point of view.

The paper is organized as follows. In Section II we briefly recall the basic definitions of β-VaR and β-CVaR and we show how the portfolio optimization problem can be reduced to linear programming. We introduce a "cost function" to be minimized under linear constraints and we discuss the rationale for a statistical mechanics approach. In Section III we solve the problem of optimizing large portfolios under ES using the replica approach. Our results and the comparison with numerics are reported in Section IV, and our conclusions are summarized in Section V.

II. THE OPTIMIZATION PROBLEM

We consider a portfolio of N financial instruments $\mathbf{w} = \{w_1, \dots w_N\}$, where w_i is the position of asset i. The global budget constraint fixes the sum of these numbers: we impose for example

$$\sum_{i=1}^{N} w_i = N \ . \tag{1}$$

We do not stipulate any constraint on short selling, so that w_i can be any negative or positive number. This is, of course, unrealistic for liquidity reasons, but considering this case allows us to show up the essence of the phenomenon.

For Peer Review Only

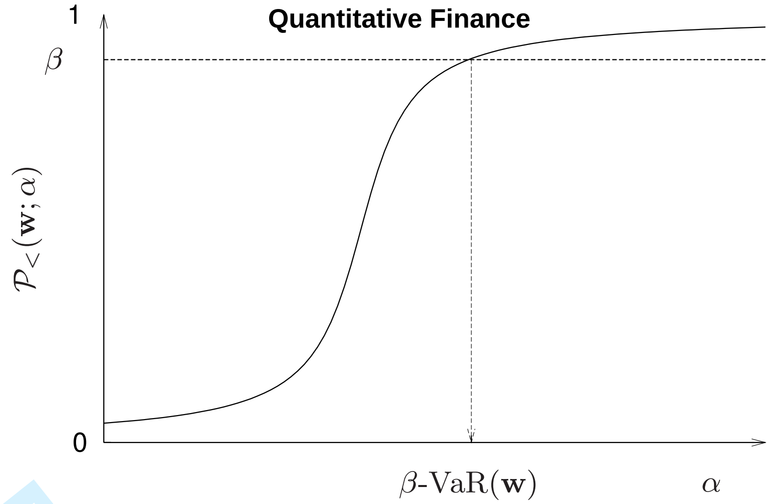

FIG. 1: Schematic representation of the VaR measure of risk. $\mathcal{P}_<(\mathbf{w})$ is the probability of a loss associated to the portfolio $\mathbf{w}$ being smaller than α. The conditional VaR β-CVaR (or ES) is the average loss when this is constrained to be greater than the β-VaR.

For Peer Review Only

If we imposed a constraint that would render the domain of the w_i bounded (such as a ban on short selling, e.g.), this would evidently prevent the weights from diverging, but a vestige of the transition would still remain in the form of large, though finite, fluctuations of the weights, and in a large number of them sticking to the "walls" of the domain.

We denote the returns on the assets by $\mathbf{x} = \{x_1, x_2, \ldots x_N\}$, and we assume that there exists an underlying probability distribution function $p(\mathbf{x})$ of the returns. The loss of portfolio $\mathbf{w}$ given the returns $\mathbf{x}$ is $\ell(\mathbf{w}|\mathbf{x}) = -\sum_{i=1}^N w_i x_i$, and the probability of that loss being smaller than a given threshold α is

$$\mathcal{P}_<(\mathbf{w},\alpha) = \int d\mathbf{x}\; p(\mathbf{x})\theta\big(\alpha - \ell(\mathbf{w}|\mathbf{x})\big) \;, \tag{2}$$

where $\theta(\cdot)$ is the Heaviside step function, equal to 1 if its argument is positive and 0 otherwise. The β-VaR of this portfolio is formally defined by

$$\beta\text{-VaR}(\mathbf{w}) = \min\{\alpha :\; \mathcal{P}_<(\mathbf{w},\alpha) \geq \beta\} \;, \tag{3}$$

(see Fig. 1), while the CVaR (or ES, in this case) associated with the same portfolio is the average loss on the tail of the distribution,

$$\beta\text{-CVaR}(\mathbf{w}) = \frac{\displaystyle\int d\mathbf{x}\; p(\mathbf{x})\ell(\mathbf{w}|\mathbf{x})\theta\big(\ell(\mathbf{w}|\mathbf{x}) - \beta\text{-VaR}(\mathbf{w})\big)}{\displaystyle\int d\mathbf{x}\; p(\mathbf{x})\theta\big(\ell(\mathbf{w}|\mathbf{x}) - \beta\text{-VaR}(\mathbf{w})\big)} = \frac{1}{1-\beta}\int d\mathbf{x}\; p(\mathbf{x})\ell(\mathbf{w}|\mathbf{x})\theta\big(\ell(\mathbf{w}|\mathbf{x}) - \beta\text{-VaR}(\mathbf{w})\big) \;. \tag{4}$$

The threshold β then represents a confidence level. In practice, the typical values of β which one considers are $\beta = 0.90, 0.95$, and 0.99, but we will address the problem for any $\beta \in [0,1]$. What is usually called "exceeding probability" in previous literature would correspond here to $(1-\beta)$.

As mentioned in the introduction, the ES measure can be obtained from a variational principle [5]. The minimization of a properly chosen objective function leads directly to (4):

$$\beta\text{-CVaR}(\mathbf{w}) \;=\; \min_v F_\beta(\mathbf{w}, v) \;, \tag{5}$$

$$F_\beta(\mathbf{w}, v) \;\equiv\; v + (1-\beta)^{-1}\int d\mathbf{x}\; p(\mathbf{x})\big[\ell(\mathbf{w}|\mathbf{x}) - v\big]^+ \;. \tag{6}$$

Here, $[a]^+ \equiv (a + |a|)/2$. The external parameter v over which one has to minimize is claimed to be relevant in itself [5], since its optimal value may represent a good estimate for the actual value-at-risk of the portfolio. We will come back to this point as we discuss our results. We stress here that minimizing (6) over $\mathbf{w}$ *and* v is equivalent to optimizing (4) over the portfolio vectors $\mathbf{w}$.

Of course, in practical cases the probability distribution of the loss is not known and must be inferred from the past data. In other words, we need an "in-sample" estimate of the integral in (6), which would turn a well posed (but useless) optimization problem into a practical approach. We thus approximate the integral by sampling the probability distributions of returns. If we have a time series $\mathbf{x}^{(1)}, \ldots \mathbf{x}^{(T)}$, our objective function becomes simply

$$\hat{F}_\beta(\mathbf{w}, v) = v + \frac{1}{(1-\beta)T} \sum_{\tau=1}^{T} \left[\ell(\mathbf{w}|\mathbf{x}^{(\tau)}) - v\right]^+ = v + \frac{1}{(1-\beta)T} \sum_{\tau=1}^{T} \left[-v - \sum_{i=1}^{N} w_i x_{i\tau}\right]^+ , \tag{7}$$

where we denote by $x_{i\tau}$ the return of asset i at time τ.

Minimizing this risk measure is the same as the following linear programming problem:

- given one data sample, i.e. a matrix $x_{i\tau}$, $i = 1, \ldots N$, $\tau = 1, \ldots T$,
- minimize the *cost function*

$$E_\beta\left[\mathbf{Y}; \{x_{i\tau}\}\right] = E_\beta\left[v, \{w_i\}, \{u_\tau\}; \{x_{i\tau}\}\right] = (1-\beta)Tv + \sum_{t=\tau}^{T} u_\tau , \tag{8}$$

- over the $(N + T + 1)$ variables $\mathbf{Y} \equiv \{w_1, \ldots w_N\, u_1, \ldots u_T\, v,$
- under the $(2T + 1)$ constraints

$$u_\tau \geq 0 , \quad u_\tau + v + \sum_{i=1}^{N} x_{i\tau} w_i \geq 0 \quad \forall \tau , \quad \text{and} \quad \sum_{i=1}^{N} w_i = N . \tag{9}$$

Since we allow short positions, not all the w_i are positive, which makes this problem different from standard linear programming. To keep the problem tractable, we impose the condition that $w_i \geq -W$, where W is a very large cutoff, and the optimization problem will be said to be ill-defined if its solution does not converge to a finite limit when $W \to \infty$. It is now clear why constraining all the w_i to be non-negative would eliminate the feasibility problem: a finite solution will always exist because the weights are by definition bounded, the worst case being an optimal portfolio with only one non-zero weight taking care of the total budget. The control parameters that govern the problem are the threshold β and the ratio N/T of assets to data points. The resulting "phase diagram" is then a line in the β-N/T plane separating a region in which, with high probability, the minimization problem is not bounded and thus does not admit a finite solution, and another region in which a finite solution exists with high probability. These statements are non-deterministic because of the intrinsic probabilistic nature of the returns. We will address this minimization problem in the non-trivial limit where $T \to \infty$, $N \to \infty$, while N/T stays finite. In this "thermodynamic" limit, we shall assume that extensive quantities (like the average loss of the optimal portfolio, i.e. the minimum cost function) do not fluctuate, namely that their probability distribution is concentrated around the mean value. This "self-averaging" property has been proved for a wide range of similar statistical mechanics models [7]. Then, we will be interested in the average value of the minimum of the cost function 8) over the distribution of returns. Given the similarity of portfolio optimization with the statistical physics of disordered systems, this problem can be addressed analytically by means of a replica approach [8].

III. THE REPLICA APPROACH

For a given history of returns x_{it}, one can compute the minimum of the cost function, $\min_{\mathbf{Y}} E_\beta[\mathbf{Y}; \{x_{it}\}]$. In this section we show how to compute analytically the expectation value of this quantity over the histories of return. For simplicity we shall keep to the case in which the x_{it} are independent identically distributed (iid) normal variables, so that an history of returns x_{it} is drawn from the distribution

$$p(\{x_{it}\}) \sim \prod_{it} e^{-N x_{it}^2/2} . \tag{10}$$

This assumption of iid normal distribution of returns is very restrictive, but we would like to emphasize that the method that we use can be generalized easily to iid variables with other distributions, and also in some cases to

For Peer Review Only

correlated variables. Certainly the precise location of the critical value of N/T separating an unfeasible phase from a feasible one depends on the distribution of returns. But we expect that the broad features like the existence of this critical value, or the way the fluctuations in portfolio diverge when approaching the transition, should not depend on this distribution. This property, called universality, has been one of the major discoveries of statistical mechanics in the last fifty years.

Instead of focusing only on the the minimal cost, the statistical mechanics approach makes a detour: it considers, for a given history of returnsx_{it}, a probability distribution in the space of variables $\mathbf{Y}$, defined by $P_\gamma(\mathbf{Y}) = 1/Z_\gamma[\{x_{it}\}]\ \exp\Big[-\gamma E_\beta[\mathbf{Y};\{x_{it}\}]\Big]$. The parameter γ is an auxiliary parameter: in physics it is the inverse of the temperature, in the present case it is just one parameter that we introduce in order to have a probability distribution on $\mathbf{Y}$ which interpolates between the uniform probability ($\gamma = 0$), and a probability which is peaked on the value of $\mathbf{Y}$ which minimizes the cost $E_\beta[\mathbf{Y};\{x_{it}\}]$ (case $\gamma = \infty$).

The normalization constant $Z_\gamma[\{x_{it}\}]$ is called the partition function at inverse temperature γ; it is defined as

$$Z_\gamma[\{x_{it}\}] \ = \ \int_V d\mathbf{Y} \exp\Big[-\gamma E_\beta[\mathbf{Y};\{x_{it}\}]\Big] \tag{11}$$

where V is the convex polytope defined by (9).

The partition function contains a lot of information on the problem. For instance the minimal cost can be expressed as $\lim_{\gamma\to\infty}(-1)/(N\gamma)\log Z_\gamma[\{x_{it}\}]$. We shall be interested in computing the large N limit of the minimal cost per variable:

$$\varepsilon[\{x_{it}\}] = \lim_{N\to\infty} \frac{\min E[\{x_{it}\}]}{N} = \lim_{N\to\infty}\lim_{\gamma\to\infty} \frac{-1}{N\gamma}\log Z_\gamma[\{x_{it}\}]\ . \tag{12}$$

In the following, we will compute the average value of this quantity over the choice of the sample x_{it}. Using equation (12), we can compute this average minimum cost if we are able to compute the average of the *logarithm* of Z. This is a difficult problem which is usually circumvented by means of the so called "replica trick": one computes the average of Z^n, where n is an integer, and then the average of the logarithm is obtained by

$$\overline{\log Z} = \lim_{n\to 0}\frac{\partial \overline{Z^n}}{\partial n}\ , \tag{13}$$

thus assuming that Z^n can be analytically continued to real values of n. The overline stands for an average over different samples, i.e. over the probability distribution (10). This technique has a long history in the physics of spin glasses [8]: the proof that it leads to the correct solution has been obtained [9] recently.

The partition function (11) can be written more explicitly as

$$\begin{aligned} Z_\gamma[\{x_{it}\}] \ &= \ \int_{-\infty}^{+\infty} dv \int_0^{+\infty}\prod_{t=1}^T du_t \int_{-\infty}^{+\infty}\prod_{i=1}^N dw_i \int_{-i\infty}^{+i\infty} d\lambda \exp\left[\lambda\left(\sum_{i=1}^N w_i - N\right)\right] \times \\ &\times \int_0^{+\infty}\prod_{t=1}^T d\mu_t \int_{-i\infty}^{+i\infty}\prod_{t=1}^T d\hat\mu_t \exp\left[\sum_{t=1}^T \hat\mu_t\left(u_t + v + \sum_{i=1}^N x_{it}w_i - \mu_t\right)\right]\exp\left[-\gamma(1-\beta)Tv - \gamma\sum_{t=1}^T u_t\right] \end{aligned} \tag{14}$$

where the constraints are imposed by means of the Lagrange multipliers $\lambda, \mu, \hat\mu$. The replica trick is based on the idea that the n-th power of the partition function appearing in (13), can be written as the partition function for n independent replicas $\mathbf{Y^1}$,...,$\mathbf{Y^n}$ of the system: all the replicas correspond to the *same* history of returns $\{x_{it}\}$, and their joint probability distribution function is $P_\gamma(\mathbf{Y^1},\ldots,\mathbf{Y^n}) = 1/Z_\gamma^n[\{x_{it}\}]\ \exp\Big[-\gamma\sum_{a=1}^n E_\beta[\mathbf{Y^a};\{x_{it}\}]\Big]$. It is not difficult to write down the expression for Z^n and average it over the distribution of samples x_{it}. One introduces the *overlap* matrix

$$Q^{ab} = \frac{1}{N}\sum_{i=1}^N w_i^a w_i^b\ , \quad a, b = 1,\ldots n\ , \tag{15}$$

as well as its conjugate $\hat Q^{ab}$ (the Lagrange multiplier imposing (15)), where a and b are replica indexes. This matrix characterizes how the portfolios in different replicas differ: they provide some indication of how the measure $P_\gamma(\mathbf{Y})$

For Peer Review Only

is spread. After (several) Gaussian integrations, one obtains

$$\overline{Z_\gamma^n[\{x_{it}\}]} \sim \int_{-\infty}^{+\infty} \prod_{a=1}^n dv^a \int_{-\infty}^{+\infty} \prod_{a,b} dQ^{ab} \int_{-i\infty}^{+i\infty} \prod_{a,b} d\hat{Q}^{ab} \exp\left\{ N \sum_{a,b} Q^{ab}\hat{Q}^{ab} - N \sum_{a,b} \hat{Q}^{ab} - \gamma(1-\beta)T \sum_a v^a \right.$$
$$\left. -Tn\log\gamma + T\log\hat{Z}_\gamma\left(\{v^a\},\{Q^{ab}\}\right) - \frac{T}{2}\mathrm{Tr}\log Q - \frac{N}{2}\mathrm{Tr}\log\hat{Q} - \frac{nN}{2}\log 2 \right\} , \tag{16}$$

where

$$\hat{Z}_\gamma\left(\{v^a\},\{Q^{ab}\}\right) \equiv \int_{-\infty}^{+\infty} \prod_{a=1}^n dy^a \exp\left[-\frac{1}{2}\sum_{a,b=1}^n (Q^{-1})^{ab}(y^a - v^a)(y^b - v^b) + \gamma \sum_{a=1}^n y^a\theta(-y^a) \right] . \tag{17}$$

We now write $T = tN$ and work at fixed t while $N \to \infty$.

The most natural solution is obtained by realizing that all the replicas are identical. Given the linear character of the problem, the symmetric solution should be the correct one. The replica-symmetric solution corresponds to the *ansatz*

$$Q^{ab} = \begin{cases} q_1 & \text{if } a = b \\ q_0 & \text{if } a \neq b \end{cases} ; \quad \hat{Q}^{ab} = \begin{cases} \hat{q}_1 & \text{if } a = b \\ \hat{q}_0 & \text{if } a \neq b \end{cases} , \tag{18}$$

and $v^a = v$ for any a. As we discuss in detail in appendix A, one can show that the optimal cost function, computed from eq. (**??**), is the minimum of

$$\varepsilon(v, q_0, \Delta) = \frac{1}{2\Delta} + \Delta\left[t(1-\beta)v - \frac{q_0}{2} + \frac{t}{2\sqrt{\pi}} \int_{-\infty}^{+\infty} ds e^{-s^2} g(v + s\sqrt{2q_0}) \right] , \tag{19}$$

where $\Delta \equiv \lim_{\gamma\to\infty} \gamma\Delta q$ and the function $g(\cdot)$ is defined as

$$g(x) = \begin{cases} 0 & x \geq 0 , \\ x^2 & -1 \leq x < 0 , \\ -2x - 1 & x < -1 . \end{cases} \tag{20}$$

Note that this function and its derivative are continuous. Moreover, v and q_0 in (19) are solutions of the saddle point equations

$$1 - \beta + \frac{1}{2\sqrt{\pi}} \int ds e^{-s^2} g'(v + s\sqrt{2q_0}) = 0 , \tag{21}$$
$$-1 + \frac{t}{\sqrt{2\pi q_0}} \int ds e^{-s^2} s\, g'(v + s\sqrt{2q_0}) = 0 . \tag{22}$$

We require that the minimum of (19) occur at a finite value of Δ. In order to understand this point, we recall the meaning of Δ (see also (18)):

$$\Delta/\gamma \sim \Delta q = (q_1 - q_0) = \frac{1}{N}\sum_{i=1}^N \left(w_i^{(1)}\right)^2 - \frac{1}{N}\sum_{i=1}^N w_i^{(1)} w_i^{(2)} \sim \overline{w^2} - \overline{w}^2 , \tag{23}$$

where the superscripts (1) and (2) represent two generic replicas of the system. We then find that Δ is proportional to the fluctuations in the distribution of the w's. An infinite value of Δ would then correspond to a portfolio which is infinitely short on some particular positions and, because of the global budget constraint (1), infinitely long on some other ones.

Given (19), the existence of a solution at finite Δ translates into the following condition:

$$t(1-\beta)v - \frac{q_0}{2} + \frac{t}{2\sqrt{\pi}} \int_{-\infty}^{+\infty} ds e^{-s^2} g(v + s\sqrt{2q_0}) \geq 0 , \tag{24}$$

which defines, along with eqs. (21) and (22), our phase diagram.

For Peer Review Only

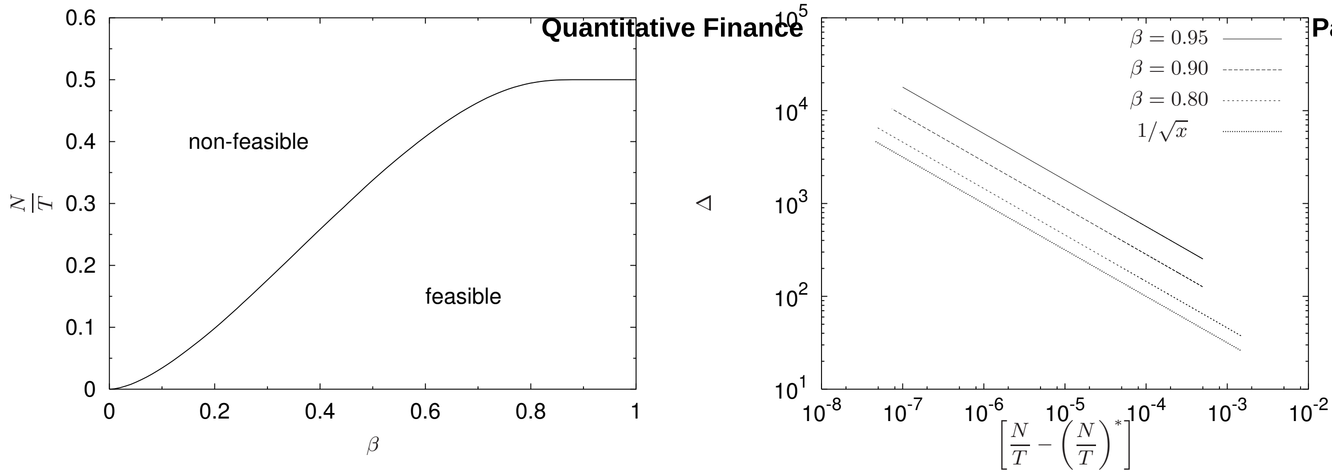

FIG. 2: Left panel: The phase diagram of the feasibility problem for expected shortfall. Right panel: The order parameter Δ diverges with an exponent 1/2 as the transition line is approached. A curve of slope $-1/2$ is also shown for comparison.

IV. THE FEASIBLE AND UNFEASIBLE REGIONS

We can now chart the feasibility map of the expected shortfall problem. Following the notation of [6], we will use as control parameters $N/T \equiv 1/t$ and β. The limiting case $\beta \to 1$ can be worked out analytically and one can show that the critical value t^* is given by

$$\frac{1}{t^*} = \frac{1}{2} - \mathcal{O}\left[(1-\beta)^3 e^{-\left(4\pi(1-\beta)^2\right)^{-1}}\right] \ . \tag{25}$$

This limit corresponds to the over-pessimistic case of maximal loss, in which the single worst loss contribute to the risk measure. The optimization problem is the following:

$$\min_{\mathbf{w}} \left[\max_t \left(-\sum_i w_i x_{it}\right)\right] \ . \tag{26}$$

A simple "geometric" argument by Kondor et al. [6] leads to the critical value $1/t^* = 0.5$ in this extreme case. The idea is the following. According to eq. (26), one has to look for the minimum of a polytope made by a large number of planes, whose normal vectors (the x_{it}) are drawn from a symmetric distribution. The simplex is convex, but with some probability it can be unbounded from below and then the optimization problem is ill defined. Increasing T means that the probability of this event decreases, because there are more planes and thus it is more likely that for large values of the w_i the max over t has a positive slope in the i-th direction. The exact law for this probability can be obtained by induction on N and T [6] and, as we said, it jumps in the thermodynamic limit from 1 to 0 at $N/T = 0.5$. The example of the max-loss risk measure is also helpful because it allows to stress two aspects of the problem: 1) even for finite N and T there is a finite chance that the risk measure is unbounded from below in some samples, and 2) the phase transition occurs in the thermodynamic limit when N/T is strictly smaller than 1, i.e. much before the covariance matrix develops zero modes. The very nature of the problem is that the risk measure there is simply not bounded from below. As for the ES risk measure, the threshold value $N/T = 0.5$ can be thought of as a good approximation of the actual value for many cases of practical interest (i.e. $\beta \gtrsim 0.9$), since the corrections to this limit case are exponentially small (eq. (25)).

For finite values of β we have solved numerically eqs. (21), (22) and (24) using the following procedure. We first solve the two equations (21) and (22), which always admit a solution for (v, q_0). We then plot the l.h.s. of eq. (24) as a function of $1/t$ for a fixed value of β. This function is positive at small $1/t$ and becomes negative beyond a threshold $1/t^*$. By keeping track of $1/t^*$ (numerically obtaining via linear interpolations) for each value of β we build up the phase diagram (Fig. 2, left). This diagram is in agreement with the numerical results obtained in ref. 6. We show in the right panel of Fig. 2 the divergence of the order parameter Δ versus $1/t - 1/t^*$. The critical exponent is found to be 1/2:

$$\Delta \sim \left(\frac{1}{t} - \frac{1}{t^*(\beta)}\right)^{-1/2} \ , \tag{27}$$

For Peer Review Only

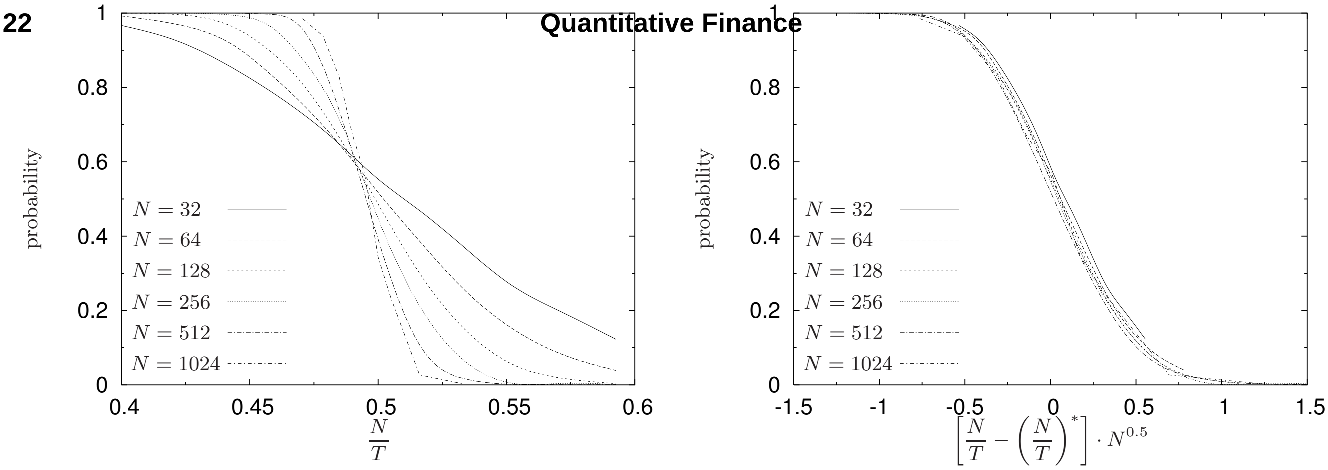

FIG. 3: Left: The probability of finding a finite solution as obtained from linear programming at increasing values of N and with $\beta = 0.8$. Right: Scaling plot of the same data. The critical value is set equal to the analytical one, $N/T = 0.4945$ and the critical exponent is $1/2$, i.e. the one obtained by Kondor et al. [6] for the limit case $\beta \to 1$. The data do not collapse perfectly, and better results can be obtained by slightly changing either the critical value or the exponent.

again in agreement with the scaling found in ref. 6. We have performed extensive numerical simulations in order to check the validity of our analytical findings. For a given realization of the time series, we solve the optimization problem (8) by standard linear programming [10]. We impose a large negative cutoff for the w's, that is $w_i > -W$, and we say that a feasible solution exists if it stays finite for $W \to \infty$. We then repeat the procedure for a certain number of samples, and then average our final results (optimal cost, optimal v, and the variance of the w's in the optimal portfolio) over those of them which produced a finite solution. In Fig. 3 we show how the probability of finding a finite solution depends on the size of the problem. Here, the probability is simply defined in terms of the frequency. We see that the convergence towards the expected 1-0 law is fairly slow, and a finite size scaling analysis is shown in the right panel. Without loss of generality, we can summarize the finite-N numerical results by writing the probability of finding a finite solution as

$$p(N, T, \beta) = f\left[\left(\frac{1}{t} - \frac{1}{t^*(\beta)}\right) \cdot N^{\alpha(\beta)}\right] , \tag{28}$$

where $f(x) \to 1$ if $x \gg 1$ and $f(x) \to 0$ if $x \ll 1$, and where $\alpha(1) = 1/2$. It is interesting to note that these results do not depend on some initial conditions of the algorithm we used to solve the problem: for a given sample, the algorithm finds in a linear time the minimum of the polytope by looking at all its vertexes exhaustively. The statistics is taken by repeating such a deterministic procedure on a large number of samples chosen at random.

In Fig. 4 (left panel) we plot, for a given value of β, the optimal cost found numerically for several values of the size N compared to the analytical prediction at infinite N. One can show that the cost vanishes as $\Delta^{-1} \sim (1/t - 1/t^*)^{1/2}$. The right panel of the same figure shows the behavior of the value of v which leads to the optimal cost versus N/T, for the same fixed value of β. Also in this case, the analytical ($N \to \infty$ limit) is plotted for comparison. We note that this quantity was suggested [5] to be a good approximation of the VaR of the optimal portfolio: We find here that v_{opt} diverges at the critical threshold and becomes negative at an even smaller value of N/T.

V. CONCLUSIONS

We have shown that the problem of optimizing a portfolio under the expected shortfall measure of risk by using empirical distributions of returns is not well defined when the ratio N/T of assets to data points is larger than a certain critical value. This value depends on the threshold β of the risk measure in a continuous way and this defines a phase diagram. The lower the value of β, the larger the length of the time series needed for the portfolio optimization. The analytical approach we have discussed in this paper allows us to have a clear understanding of this phase transition. The mathematical reason for the non-feasibility of the optimization problem is that, with a certain probability $p(N, T, \beta)$, the linear constraints in (9) define a simplex which is not bounded from below, thus leading to a solution which is not finite ($\Delta q \to \infty$ in our language), in the same way as it happens in the extreme case $\beta \to 1$

For Peer Review Only

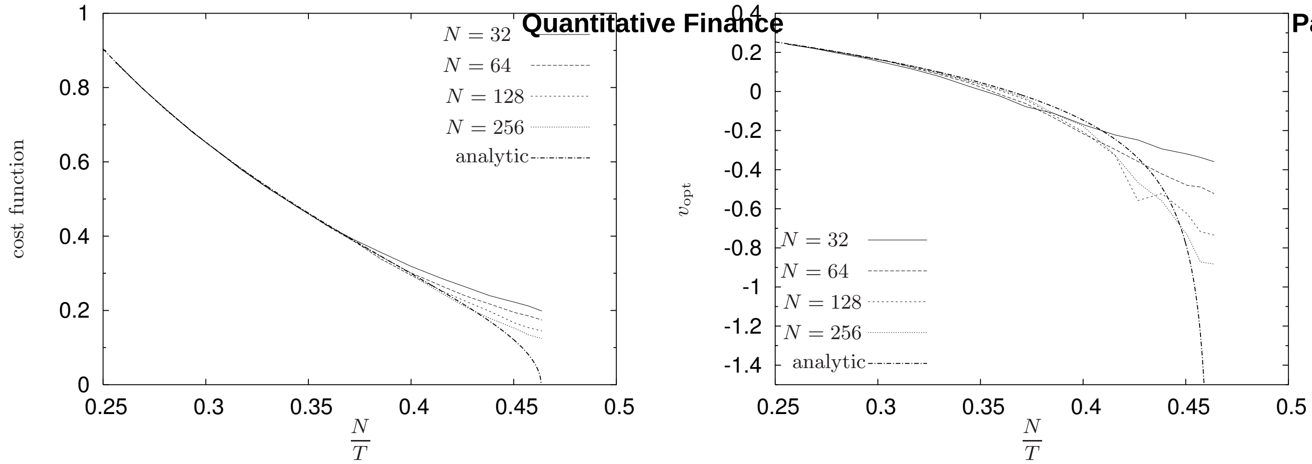

FIG. 4: Numerical results from linear programming and comparison with analytical predictions at large N. Left: The minimum cost of the optimization problem vs N/T, at increasing values of N. The thick line is the analytical solution (19). Here $\beta = 0.7$, $(N/T)^* \simeq 0.463$. Right: The optimal value of v as found numerically for several values of N is compared to the analytical solution.

discussed in [6]. From a more physical point of view, it is reasonable that the feasibility of the problem depend on the number of data points we take from the time series with respect to the number of financial instruments of our portfolio. The probabilistic character of the time series is reflected in the probability $p(N, T, \beta)$. Interestingly, this probability becomes a threshold function at large N if $N/T \equiv 1/t$ is finite, and its general form is given in (28).

These results have a practical relevance in portfolio optimization. The order parameter discussed in this paper is tightly related to the relative estimation error [6]. The fact that this order parameter has been found to diverge means that in some regions of the parameter space the estimation error blows up which makes the task of portfolio optimization completely meaningless. The divergence of estimation error is not limited to the case of expected shortfall: as shown in [6], it happens in the case of variance and absolute deviation as well, but the noise sensitivity of expected shortfall turns out to be even greater than that of these more conventional risk measures.

There is nothing surprising about the fact that if there are no sufficient data, the estimation error is large and we cannot make a good decision. What is surprising is the fact that there is a sharply defined threshold where the estimation error actually diverges.

For a given portfolio size, it is important to know that a minimum amount of data points is required in order to perform an optimization based on empirical distributions. We also note that the divergence of the parameter Δ at the phase transition, which is directly related to the fluctuations of the optimal portfolio, may play a dramatic role in practical cases. To stress this point, we can define a sort of "susceptibility" with respect to the data,

$$\chi_{ij}^{t} = \frac{\partial \langle w_j \rangle}{\partial x_{it}} \ , \tag{29}$$

and one can show that this quantity diverges at the critical point, since $\chi_{ij} \sim \Delta$. A small change (or uncertainty) in x_{it} becomes increasingly relevant as the transition is approached, and the portfolio optimization could then be very unstable even in the feasible region of the phase diagram. We stress that the susceptibility we have introduced might be considered as a measure of the effect of the noise on portfolio selection and is very reminiscent of the measure proposed in [11].

In order to present a clean, analytic picture, we have made several simplifying assumptions in this work. We have omitted the constraint on the returns, liquidity constraints, correlations between the assets, non-stationary effects, etc. Some of these can be systematically taken into account and we plan to return to these finer details in a subsequent work.

Acknowledgments. We thank O. C. Martin, and M. Potters for useful discussions, and particularly J. P. Bouchaud for a critical reading of the manuscript. S. C. is supported by EC through the network MTR 2002-00319, STIPCO, I.K. by the National Office of Research and Technology under grant No. KCKHA005.

For Peer Review Only

APPENDIX A: THE REPLICA SYMMETRIC SOLUTION

We show in this appendix how the minimum cost function corresponding to the replica-symmetric ansatz is obtained.

The 'TrLog$Q$' term in (16) is computed by realizing that the eigenvalues of such a symmetric matrix are $(q_1 + (n-1)q_0)$ (with multiplicity 1) and $(q_1 - q_0)$ with multiplicity $n-1$. Then,

$$\mathrm{Tr}\log Q = \log\det Q = \log(q_1 + (n-1)q_0) + (n-1)\log(q_1 - q_0) = n\left(\log \Delta q + \frac{q_1}{\Delta q}\right) + \mathcal{O}(n^2)\ , \tag{A1}$$

where $\Delta q \equiv q_1 - q_0$. The effective partition function in (17) depends on Q^{-1}, whose elements are:

$$(Q^{-1})^{ab} = \begin{cases} (\Delta q - q_0)/(\Delta q)^2 + \mathcal{O}(n) & \text{if } a = b \\ -q_0/(\Delta q)^2 + \mathcal{O}(n) & \text{if } a \neq b \end{cases} \tag{A2}$$

By introducing a Gaussian measure $dP_{q_0}(s) \equiv \frac{ds}{\sqrt{2\pi q_0}} e^{-s^2/2q_0}$, one can show that

$$\begin{aligned} \frac{1}{n}\log \hat{Z}(v, q_1, q_0) &= \frac{1}{n}\log\left\{\int \prod_a dx_a e^{-\frac{1}{2\Delta q}\sum_a (x^a)^2 + \gamma\sum_a (x^a+v)\theta(-x^a - v)} \int dP_{q_0}(s) e^{\frac{s}{\Delta q}\sum_a x^a}\right\} \\ &= \frac{q_0}{2\Delta q} + \int dP_{q_0}(s)\log B_\gamma(s, v, \Delta q) + \mathcal{O}(n) \end{aligned} \tag{A3}$$

where we have defined

$$B_\gamma(s, v, \Delta q) \equiv \int dx \exp\left(-\frac{(x-s)^2}{2\Delta q} + \gamma(x+v)\theta(-x-v)\right)\ . \tag{A4}$$

The exponential in (16) now reads $\exp Nn[S(q_0, \Delta q, \hat{q}_0, \Delta\hat{q}) + \mathcal{O}(n)]$, where

$$\begin{aligned} S(q_0, \Delta q, \hat{q}_0, \Delta\hat{q}) = q_0\Delta\hat{q} + \hat{q}_0\Delta q + \Delta q\Delta\hat{q} - \Delta\hat{q} - \gamma t(1-\beta)v - t\log\gamma + t\int dP_{q_0}(s)\log B_\gamma(s, v, \Delta q) \\ -\frac{t}{2}\log\Delta q - \frac{1}{2}\left(\log\Delta\hat{q} + \frac{\hat{q}_0}{\Delta\hat{q}}\right) - \frac{\log 2}{2}\ . \end{aligned} \tag{A5}$$

The saddle point equations for $\hat{q}_0$ and $\Delta\hat{q}$ allow then to simplify this expression. The free energy $(-\gamma)f_\gamma = \lim_{n\to 0}\partial\overline{Z_\gamma^n}/\partial n$ is given by

$$-\gamma f_\gamma(v, q_0, \Delta q) = \frac{1}{2} - t\log\gamma + \frac{1-t}{2}\log\Delta q + \frac{q_0 - 1}{2\Delta q} - \gamma t(1-\beta)v + t\int dP_{q_0}(s)\log B_\gamma(s, v, \Delta q)\ , \tag{A6}$$

where the actual values of v, q_0 and Δq are fixed by the saddle point equations

$$\frac{\partial f_\gamma}{\partial v} = \frac{\partial f_\gamma}{\partial q_0} = \frac{\partial f_\gamma}{\partial \Delta q} = 0\ . \tag{A7}$$

A close inspection of these saddle point equations allows one to perform the low temperature $\gamma \to \infty$ limit by assuming that $\Delta q = \Delta/\gamma$ while v and q_0 do not depend on the temperature. In this limit one can show that

$$\lim_{\gamma\to\infty}\frac{1}{\gamma}\log B_\gamma(s, v, \Delta/\gamma) = \begin{cases} s + v + \Delta/2 & s < -v - \Delta \\ -(v+s)^2/2\Delta & -v - \Delta \le s < -v \\ 0 & s \ge -v \end{cases} \tag{A8}$$

If we plug this expression into eq. (A6) and perform the large-γ limit we get the minimum cost:

$$E = \lim_{\gamma\to\infty} f_\gamma = -\frac{q_0 - 1}{2\Delta} + t(1-\beta)v - t\int_{-\infty}^{-\Delta}\frac{dx}{\sqrt{2\pi q_0}}e^{-\frac{(x-v)^2}{2q_0}}\left(x + \frac{\Delta}{2}\right) + \frac{t}{2\Delta}\int_{-\Delta}^{0}\frac{dx}{\sqrt{2\pi q_0}}e^{-\frac{(x-v)^2}{2q_0}}x^2\ . \tag{A9}$$

We rescale $x \to x\Delta$, $v \to v\Delta$, and $q_0 \to q_0\Delta^2$, and after some algebra we obtain eq. (19).

---

[1] see P. Artzner, F. Delbaen, J. M. Eber, and D. Heath, Mathematical Finance **9,** 203–228 (1999), for an axiomatic definition of coherence.

For Peer Review Only

[2] C. Acerbi, and D. Tasche, Journal of Banking and Finance **26,** 1487–1503 (2002).
[3] R. Frey, and A. J. McNeil, (2002).
[4] G. C. Pflug, in "Probabilistic Constrained Optimization: Methodology and Applications", S. Uryasev (ed.), Kluwer Academic Publisher (2000).
[5] R. Rockafellar, and S. Uryasev, The Journal of Risk **2,** 21–41 (2000).
[6] I. Kondor, Noise sensitivity of portfolio selection under various risk measures, Lectures given at the Summer School on "Risk Measurement and Control", Rome, June 9-17, 2005; I. Kondor, S. Pafka, and G. Nagy, Noise sensitivity of portfolio selection under various risk measures, submitted to J. of Banking and Finance.
[7] see e.g. F. Guerra, F.L. Toninelli, J. Stat. Phys. **115,** 531 (2004) and references therein.
[8] M. Mézard, G. Parisi, and M. .A. .Virasoro, "Spin Glass theroy and Beyond", World Scientific Lecture Notes in Physics Vol. 9, Singapore (1987).
[9] M. Talagrand, *Spin Glasses: a Challenge for Mathematicians*, Spinger-Verlag (2002).
[10] W. H. Press, S. H. Teukolsky, W. T. Wetterling, and B. P. Flannery, "Numerical Recipes in C", Cambridge University Press (Cambridge, UK, 1992).
[11] S. Pafka, and I. Kondor, Europ. Phys. J. **B27,** 277 (2002).
[12] see [2] for the subtleties related to a discrete distribution

For Peer Review Only

QUANTITATIVE FINANCE

On the Feasibility of Portfolio Optimization under Expected Shortfall

Journal:	*Quantitative Finance*
Manuscript ID:	RQUF-2006-0108.R1
Manuscript Category:	Research Paper
Date Submitted by the Author:	28-Jan-2007
Complete List of Authors:	ciliberti, stefano; univ paris sud, LPTMS Kondor, Imre Mezard, Marc
Keywords:	Portfolio Optimization, Risk Measures, Statistical Physics, Financial Applications
JEL Code:	Portfolio Optimization, Risk Measures, Statistical Physics, Financial Applications
Note: The following files were submitted by the author for peer review, but cannot be converted to PDF. You must view these files (e.g. movies) online.	
resubmission.tgz	

scholarONE™ Manuscript Central

For Peer Review Only

On the Feasibility of Portfolio Optimization under Expected Shortfall

Stefano Ciliberti,[1] Imre Kondor,[2] and Marc Mézard[1]
[1] CNRS; Univ. Paris Sud, UMR8626, LPTMS, ORSAY CEDEX, F-91405
[2] Collegium Budapest, 1014 Budapest, Szenthromsg u. 2

We address the problem of portfolio optimization under the simplest coherent risk measure, i.e. the expected shortfall. As it is well known, one can map this problem into a linear programming setting. For some values of the external parameters, when the available time series is too short, the portfolio optimization is ill posed because it leads to unbounded positions, infinitely short on some assets and infinitely long on some others. As first observed by Kondor and coworkers, this phenomenon is actually a phase transition. We investigate the nature of this transition by means of a replica approach.

For Peer Review Only

I. INTRODUCTION

Among the several risk measures existing in the context of portfolio optimization, expected shortfall (ES) has certainly gained increasing popularity in recent years. In several practical applications, ES is starting to replace the classical Value-at-Risk (VaR). There are a number of reasons for this. For a given threshold probability β, the VaR is defined so that with probability β the loss will be smaller than VaR. This definition only gives the minimum loss one can reasonably expect but does not tell anything about the typical value of that loss that can be measured by the *conditional* value-at-risk (CVaR, which is the same as ES for continuous distributions that we consider here [12]). We will be more precise on these definitions below. The point we want to stress here is that the VaR measure, lacking the mandatory properties of subadditivity and convexity,is not *coherent* [1]. This means that summing VaR's of individual portfolios will not necessarily produce an upper bound for the VaR of the combined portfolio, thus contradicting the holy principle of diversification in finance. A nice practical example of the inconsistency of VaR in credit portfolio management is reported in Ref. 3. On the other hand, it has been shown [2] that ES is a coherent measure with interesting properties [4]. Moreover, the optimization of ES can be reduced to linear programming [5] (which allows for a fast implementation) and leads to a good estimate for the VaR as a byproduct of the minimization process. To summarize, the intuitive and simple character, together with the mathematical properties (coherence) and the fast algorithmic implementation (linear programming), are the main reasons behind the growing importance of ES as a risk measure.

In this paper, we will focus on the feasibility of the portfolio optimization problem under the ES measure of risk. The control parameters of this problem are (i) the imposed threshold in probability, β, and (ii) the ratio N/T between the number N of financial assets making up the portfolio and the time series length T used to sample the probability distribution of returns. (It is curious that, albeit trivial, the scaling in N/T had not been explicitly pointed out before [11]. It has been discovered by Kondor at al. Ref. 6 that, for certain values of these parameters, the optimization problem does not have a finite solution because, even if convex, it is not bounded from below. Extended numerical simulations allowed these authors to determine the feasibility map of the problem. Here, in order to better understand the root of the problem and to study the transition from a feasible regime to an unfeasible one (corresponding to an ill-posed minimization problem) we address the same problem from an analytical point of view.

The paper is organized as follows. In Section II we briefly recall the basic definitions of β-VaR and β-CVaR and we show how the portfolio optimization problem can be reduced to linear programming. We introduce a "cost function" to be minimized under linear constraints and we discuss the rationale for a statistical mechanics approach. In Section III we solve the problem of optimizing large portfolios under ES using the replica approach. Our results and the comparison with numerics are reported in Section IV, and our conclusions are summarized in Section V.

II. THE OPTIMIZATION PROBLEM

We consider a portfolio of N financial instruments $\mathbf{w} = \{w_1, \ldots w_N\}$, where w_i is the position of asset i. The global budget constraint fixes the sum of these numbers: we impose for example

$$\sum_{i=1}^{N} w_i = N \; . \tag{1}$$

We do not stipulate any constraint on short selling, so that w_i can be any negative or positive number. This is, of course, unrealistic for liquidity reasons, but considering this case allows us to show up the essence of the phenomenon.

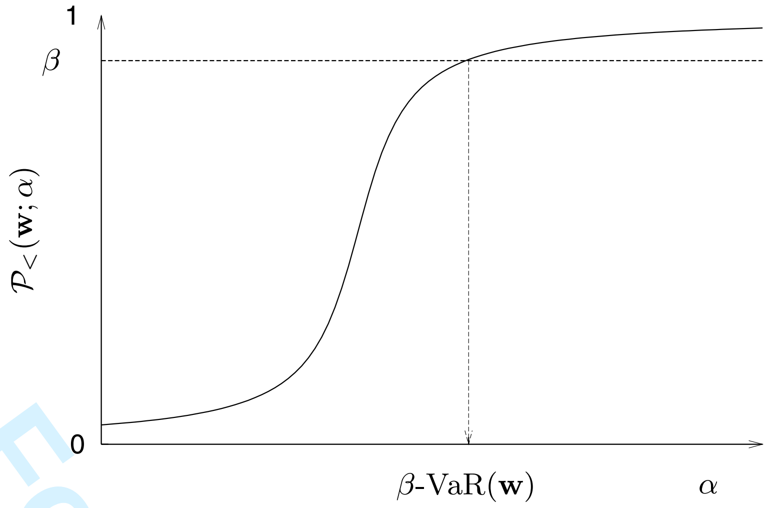

FIG. 1: Schematic representation of the VaR measure of risk. $\mathcal{P}_<(\mathbf{w})$ is the probability of a loss associated to the portfolio $\mathbf{w}$ being smaller than α. The conditional VaR β-CVaR (or ES) is the average loss when this is constrained to be greater than the β-VaR.

If we imposed a constraint that would render the domain of the w_i bounded (such as a ban on short selling, e.g.), this would evidently prevent the weights from diverging, but a vestige of the transition would still remain in the form of large, though finite, fluctuations of the weights, and in a large number of them sticking to the "walls" of the domain.

We denote the returns on the assets by $\mathbf{x} = \{x_1, x_2, \ldots x_N\}$, and we assume that there exists an underlying probability distribution function $p(\mathbf{x})$ of the returns. The loss of portfolio $\mathbf{w}$ given the returns $\mathbf{x}$ is $\ell(\mathbf{w}|\mathbf{x}) = -\sum_{i=1}^N w_i x_i$, and the probability of that loss being smaller than a given threshold α is

$$\mathcal{P}_<(\mathbf{w}, \alpha) = \int d\mathbf{x}\; p(\mathbf{x})\theta\big(\alpha - \ell(\mathbf{w}|\mathbf{x})\big) \;, \tag{2}$$

where $\theta(\cdot)$ is the Heaviside step function, equal to 1 if its argument is positive and 0 otherwise. The β-VaR of this portfolio is formally defined by

$$\beta\text{-VaR}(\mathbf{w}) = \min\{\alpha : \; \mathcal{P}_<(\mathbf{w}, \alpha) \geq \beta\} \;, \tag{3}$$

(see Fig. 1), while the CVaR (or ES, in this case) associated with the same portfolio is the average loss on the tail of the distribution,

$$\beta\text{-CVaR}(\mathbf{w}) = \frac{\int d\mathbf{x}\; p(\mathbf{x})\ell(\mathbf{w}|\mathbf{x})\theta\big(\ell(\mathbf{w}|\mathbf{x}) - \beta\text{-VaR}(\mathbf{w})\big)}{\int d\mathbf{x}\; p(\mathbf{x})\theta\big(\ell(\mathbf{w}|\mathbf{x}) - \beta\text{-VaR}(\mathbf{w})\big)} = \frac{1}{1-\beta}\int d\mathbf{x}\; p(\mathbf{x})\ell(\mathbf{w}|\mathbf{x})\theta\big(\ell(\mathbf{w}|\mathbf{x}) - \beta\text{-VaR}(\mathbf{w})\big) \;. \tag{4}$$

The threshold β then represents a confidence level. In practice, the typical values of β which one considers are $\beta = 0.90, 0.95$, and 0.99, but we will address the problem for any $\beta \in [0, 1]$. What is usually called "exceeding probability" in previous literature would correspond here to $(1 - \beta)$.

As mentioned in the introduction, the ES measure can be obtained from a variational principle [5]. The minimization of a properly chosen objective function leads directly to (4):

$$\beta\text{-CVaR}(\mathbf{w}) \;=\; \min_v F_\beta(\mathbf{w}, v) \;, \tag{5}$$

$$F_\beta(\mathbf{w}, v) \;\equiv\; v + (1-\beta)^{-1}\int d\mathbf{x}\; p(\mathbf{x})\big[\ell(\mathbf{w}|\mathbf{x}) - v\big]^+ \;. \tag{6}$$

Here, $[a]^+ \equiv (a + |a|)/2$. The external parameter v over which one has to minimize is claimed to be relevant in itself [5], since its optimal value may represent a good estimate for the actual value-at-risk of the portfolio. We will

For Peer Review Only

come back to this point as we discuss our results. We stress here that minimizing (6) over $\mathbf{w}$ *and* v is equivalent to optimizing (4) over the portfolio vectors $\mathbf{w}$.

Of course, in practical cases the probability distribution of the loss is not known and must be inferred from the past data. In other words, we need an "in-sample" estimate of the integral in (6), which would turn a well posed (but useless) optimization problem into a practical approach. We thus approximate the integral by sampling the probability distributions of returns. If we have a time series $\mathbf{x}^{(1)}, \ldots \mathbf{x}^{(T)}$, our objective function becomes simply

$$\hat{F}_\beta(\mathbf{w}, v) = v + \frac{1}{(1-\beta)T} \sum_{\tau=1}^{T} \left[\ell(\mathbf{w}|\mathbf{x}^{(\tau)}) - v\right]^+ = v + \frac{1}{(1-\beta)T} \sum_{\tau=1}^{T} \left[-v - \sum_{i=1}^{N} w_i x_{i\tau}\right]^+ , \tag{7}$$

where we denote by $x_{i\tau}$ the return of asset i at time τ.

Minimizing this risk measure is the same as the following linear programming problem:

- given one data sample, i.e. a matrix $x_{i\tau}$, $i = 1, \ldots N$, $\tau = 1, \ldots T$,
- minimize the *cost function*

$$E_\beta\left[\mathbf{Y}; \{x_{i\tau}\}\right] = E_\beta\left[v, \{w_i\}, \{u_\tau\}; \{x_{i\tau}\}\right] = (1-\beta)Tv + \sum_{t=\tau}^{T} u_\tau , \tag{8}$$

- over the $(N + T + 1)$ variables $\mathbf{Y} \equiv \{w_1, \ldots w_N\, u_1, \ldots u_T\, v,$
- under the $(2T + 1)$ constraints

$$u_\tau \geq 0 , \quad u_\tau + v + \sum_{i=1}^{N} x_{i\tau} w_i \geq 0 \quad \forall \tau , \quad \text{and} \quad \sum_{i=1}^{N} w_i = N . \tag{9}$$

Since we allow short positions, not all the w_i are positive, which makes this problem different from standard linear programming. To keep the problem tractable, we impose the condition that $w_i \geq -W$, where W is a very large cutoff, and the optimization problem will be said to be ill-defined if its solution does not converge to a finite limit when $W \to \infty$. It is now clear why constraining all the w_i to be non-negative would eliminate the feasibility problem: a finite solution will always exist because the weights are by definition bounded, the worst case being an optimal portfolio with only one non-zero weight taking care of the total budget. The control parameters that govern the problem are the threshold β and the ratio N/T of assets to data points. The resulting "phase diagram" is then a line in the β-N/T plane separating a region in which, with high probability, the minimization problem is not bounded and thus does not admit a finite solution, and another region in which a finite solution exists with high probability. These statements are non-deterministic because of the intrinsic probabilistic nature of the returns. We will address this minimization problem in the non-trivial limit where $T \to \infty$, $N \to \infty$, while N/T stays finite. In this "thermodynamic" limit, we shall assume that extensive quantities (like the average loss of the optimal portfolio, i.e. the minimum cost function) do not fluctuate, namely that their probability distribution is concentrated around the mean value. This "self-averaging" property has been proved for a wide range of similar statistical mechanics models [7]. Then, we will be interested in the average value of the minimum of the cost function 8) over the distribution of returns. Given the similarity of portfolio optimization with the statistical physics of disordered systems, this problem can be addressed analytically by means of a replica approach [8].

III. THE REPLICA APPROACH

For a given history of returns x_{it}, one can compute the minimum of the cost function, $\min_{\mathbf{Y}} E_\beta[\mathbf{Y}; \{x_{it}\}]$. In this section we show how to compute analytically the expectation value of this quantity over the histories of return. For simplicity we shall keep to the case in which the x_{it} are independent identically distributed (iid) normal variables, so that an history of returns x_{it} is drawn from the distribution

$$p(\{x_{it}\}) \sim \prod_{it} e^{-N x_{it}^2/2} . \tag{10}$$

This assumption of iid normal distribution of returns is very restrictive, but we would like to emphasize that the method that we use can be generalized easily to iid variables with other distributions, and also in some cases to

For Peer Review Only

correlated variables. Certainly the precise location of the critical value of N/T separating an unfeasible phase from a feasible one depends on the distribution of returns. But we expect that the broad features like the existence of this critical value, or the way the fluctuations in portfolio diverge when approaching the transition, should not depend on this distribution. This property, called universality, has been one of the major discoveries of statistical mechanics in the last fifty years.

Instead of focusing only on the the minimal cost, the statistical mechanics approach makes a detour: it considers, for a given history of returnsx_{it}, a probability distribution in the space of variables $\mathbf{Y}$, defined by $P_\gamma(\mathbf{Y}) = 1/Z_\gamma[\{x_{it}\}]\ \exp\Big[-\gamma E_\beta[\mathbf{Y};\{x_{it}\}]\Big]$. The parameter γ is an auxiliary parameter: in physics it is the inverse of the temperature, in the present case it is just one parameter that we introduce in order to have a probability distribution on $\mathbf{Y}$ which interpolates between the uniform probability ($\gamma = 0$), and a probability which is peaked on the value of $\mathbf{Y}$ which minimizes the cost $E_\beta[\mathbf{Y};\{x_{it}\}]$ (case $\gamma = \infty$).

The normalization constant $Z_\gamma[\{x_{it}\}]$ is called the partition function at inverse temperature γ; it is defined as

$$Z_\gamma[\{x_{it}\}] \;=\; \int_V d\mathbf{Y} \exp\Big[-\gamma E_\beta[\mathbf{Y};\{x_{it}\}]\Big] \tag{11}$$

where V is the convex polytope defined by (9).

The partition function contains a lot of information on the problem. For instance the minimal cost can be expressed as $\lim_{\gamma\to\infty}(-1)/(N\gamma)\log Z_\gamma[\{x_{it}\}]$. We shall be interested in computing the large N limit of the minimal cost per variable:

$$\varepsilon[\{x_{it}\}] = \lim_{N\to\infty}\frac{\min E[\{x_{it}\}]}{N} = \lim_{N\to\infty}\lim_{\gamma\to\infty}\frac{-1}{N\gamma}\log Z_\gamma[\{x_{it}\}]\ . \tag{12}$$

In the following, we will compute the average value of this quantity over the choice of the sample x_{it}. Using equation (12), we can compute this average minimum cost if we are able to compute the average of the *logarithm* of Z. This is a difficult problem which is usually circumvented by means of the so called "replica trick": one computes the average of Z^n, where n is an integer, and then the average of the logarithm is obtained by

$$\overline{\log Z} = \lim_{n\to 0}\frac{\partial\overline{Z^n}}{\partial n}\ , \tag{13}$$

thus assuming that Z^n can be analytically continued to real values of n. The overline stands for an average over different samples, i.e. over the probability distribution (10). This technique has a long history in the physics of spin glasses [8]: the proof that it leads to the correct solution has been obtained [9] recently.

The partition function (11) can be written more explicitly as

$$\begin{aligned} Z_\gamma[\{x_{it}\}] \;=\; & \int_{-\infty}^{+\infty} dv \int_0^{+\infty}\prod_{t=1}^T du_t \int_{-\infty}^{+\infty}\prod_{i=1}^N dw_i \int_{-i\infty}^{+i\infty} d\lambda \exp\left[\lambda\left(\sum_{i=1}^N w_i - N\right)\right]\times \\ & \times\int_0^{+\infty}\prod_{t=1}^T d\mu_t \int_{-i\infty}^{+i\infty}\prod_{t=1}^T d\hat\mu_t \exp\left[\sum_{t=1}^T \hat\mu_t\left(u_t + v + \sum_{i=1}^N x_{it}w_i - \mu_t\right)\right]\exp\left[-\gamma(1-\beta)Tv - \gamma\sum_{t=1}^T u_t\right] \end{aligned} \tag{14}$$

where the constraints are imposed by means of the Lagrange multipliers $\lambda, \mu, \hat\mu$. The replica trick is based on the idea that the n-th power of the partition function appearing in (13), can be written as the partition function for n independent replicas $\mathbf{Y^1}$,...,$\mathbf{Y^n}$ of the system: all the replicas correspond to the *same* history of returns $\{x_{it}\}$, and their joint probability distribution function is $P_\gamma(\mathbf{Y^1},\ldots,\mathbf{Y^n}) = 1/Z_\gamma^n[\{x_{it}\}]\ \exp\Big[-\gamma\sum_{a=1}^n E_\beta[\mathbf{Y^a};\{x_{it}\}]\Big]$. It is not difficult to write down the expression for Z^n and average it over the distribution of samples x_{it}. One introduces the *overlap* matrix

$$Q^{ab} = \frac{1}{N}\sum_{i=1}^N w_i^a w_i^b\ ,\quad a,b = 1,\ldots n\ , \tag{15}$$

as well as its conjugate $\hat Q^{ab}$ (the Lagrange multiplier imposing (15)), where a and b are replica indexes. This matrix characterizes how the portfolios in different replicas differ: they provide some indication of how the measure $P_\gamma(\mathbf{Y})$

is spread. After (several) Gaussian integrations, one obtains

$$\overline{Z_\gamma^n[\{x_{it}\}]} \sim \int_{-\infty}^{+\infty}\prod_{a=1}^n dv^a \int_{-\infty}^{+\infty}\prod_{a,b} dQ^{ab}\int_{-\mathrm{i}\infty}^{+\mathrm{i}\infty}\prod_{a,b} d\hat{Q}^{ab}\exp\left\{N\sum_{a,b}Q^{ab}\hat{Q}^{ab} - N\sum_{a,b}\hat{Q}^{ab} - \gamma(1-\beta)T\sum_a v^a \right.$$
$$\left. -Tn\log\gamma + T\log\hat{Z}_\gamma\left(\{v^a\},\{Q^{ab}\}\right) - \frac{T}{2}\mathrm{Tr}\log Q - \frac{N}{2}\mathrm{Tr}\log\hat{Q} - \frac{nN}{2}\log 2\right\}, \tag{16}$$

where

$$\hat{Z}_\gamma\left(\{v^a\},\{Q^{ab}\}\right) \equiv \int_{-\infty}^{+\infty}\prod_{a=1}^n dy^a \exp\left[-\frac{1}{2}\sum_{a,b=1}^n (Q^{-1})^{ab}(y^a-v^a)(y^b-v^b) + \gamma\sum_{a=1}^n y^a\theta(-y^a)\right]. \tag{17}$$

We now write $T = tN$ and work at fixed t while $N \to \infty$.

The most natural solution is obtained by realizing that all the replicas are identical. Given the linear character of the problem, the symmetric solution should be the correct one. The replica-symmetric solution corresponds to the *ansatz*

$$Q^{ab} = \begin{cases} q_1 & \text{if } a = b \\ q_0 & \text{if } a \neq b \end{cases}; \quad \hat{Q}^{ab} = \begin{cases} \hat{q}_1 & \text{if } a = b \\ \hat{q}_0 & \text{if } a \neq b \end{cases}, \tag{18}$$

and $v^a = v$ for any a. As we discuss in detail in appendix A, one can show that the optimal cost function, computed from eq. (**??**), is the minimum of

$$\varepsilon(v, q_0, \Delta) = \frac{1}{2\Delta} + \Delta\left[t(1-\beta)v - \frac{q_0}{2} + \frac{t}{2\sqrt{\pi}}\int_{-\infty}^{+\infty} ds e^{-s^2} g(v + s\sqrt{2q_0})\right], \tag{19}$$

where $\Delta \equiv \lim_{\gamma\to\infty}\gamma\Delta q$ and the function $g(\cdot)$ is defined as

$$g(x) = \begin{cases} 0 & x \geq 0 , \\ x^2 & -1 \leq x < 0 , \\ -2x - 1 & x < -1 . \end{cases} \tag{20}$$

Note that this function and its derivative are continuous. Moreover, v and q_0 in (19) are solutions of the saddle point equations

$$1 - \beta + \frac{1}{2\sqrt{\pi}}\int ds e^{-s^2} g'(v + s\sqrt{2q_0}) = 0 , \tag{21}$$
$$-1 + \frac{t}{\sqrt{2\pi q_0}}\int ds e^{-s^2} s\, g'(v + s\sqrt{2q_0}) = 0 . \tag{22}$$

We require that the minimum of (19) occur at a finite value of Δ. In order to understand this point, we recall the meaning of Δ (see also (18)):

$$\Delta/\gamma \sim \Delta q = (q_1 - q_0) = \frac{1}{N}\sum_{i=1}^N \left(w_i^{(1)}\right)^2 - \frac{1}{N}\sum_{i=1}^N w_i^{(1)} w_i^{(2)} \sim \overline{w^2} - \overline{w}^2 , \tag{23}$$

where the superscripts (1) and (2) represent two generic replicas of the system. We then find that Δ is proportional to the fluctuations in the distribution of the w's. An infinite value of Δ would then correspond to a portfolio which is infinitely short on some particular positions and, because of the global budget constraint (1), infinitely long on some other ones.

Given (19), the existence of a solution at finite Δ translates into the following condition:

$$t(1-\beta)v - \frac{q_0}{2} + \frac{t}{2\sqrt{\pi}}\int_{-\infty}^{+\infty} ds e^{-s^2} g(v + s\sqrt{2q_0}) \geq 0 , \tag{24}$$

which defines, along with eqs. (21) and (22), our phase diagram.

For Peer Review Only

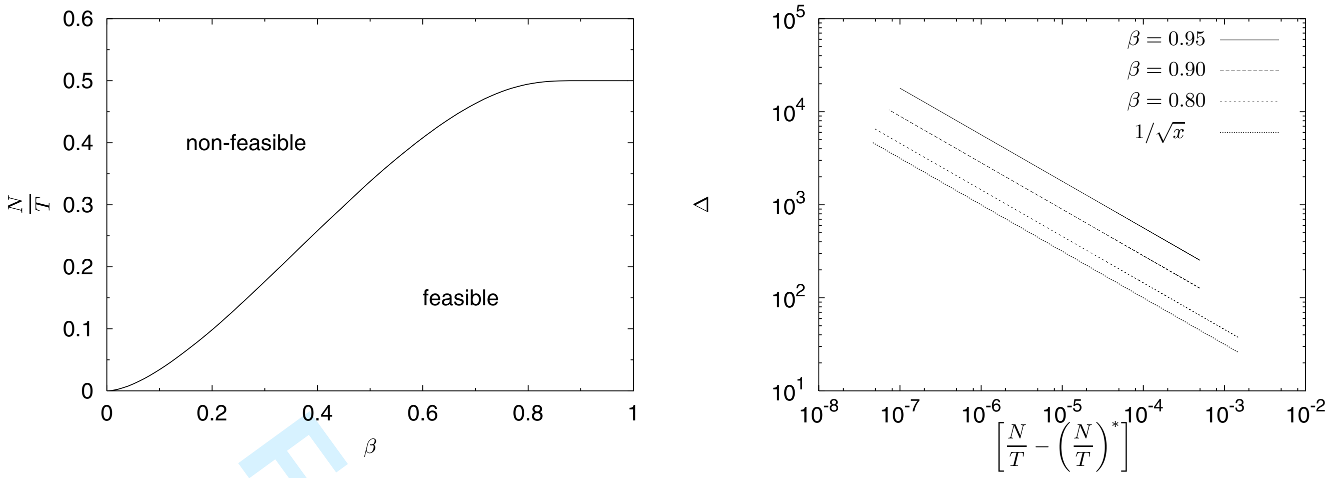

FIG. 2: Left panel: The phase diagram of the feasibility problem for expected shortfall. Right panel: The order parameter Δ diverges with an exponent 1/2 as the transition line is approached. A curve of slope $-1/2$ is also shown for comparison.

IV. THE FEASIBLE AND UNFEASIBLE REGIONS

We can now chart the feasibility map of the expected shortfall problem. Following the notation of [6], we will use as control parameters $N/T \equiv 1/t$ and β. The limiting case $\beta \to 1$ can be worked out analytically and one can show that the critical value t^* is given by

$$\frac{1}{t^*} = \frac{1}{2} - \mathcal{O}\left[(1-\beta)^3 e^{-\left(4\pi(1-\beta)^2\right)^{-1}}\right] \ . \tag{25}$$

This limit corresponds to the over-pessimistic case of maximal loss, in which the single worst loss contribute to the risk measure. The optimization problem is the following:

$$\min_{\mathbf{w}} \left[\max_t \left(-\sum_i w_i x_{it}\right)\right] \ . \tag{26}$$

A simple "geometric" argument by Kondor et al. [6] leads to the critical value $1/t^* = 0.5$ in this extreme case. The idea is the following. According to eq. (26), one has to look for the minimum of a polytope made by a large number of planes, whose normal vectors (the x_{it}) are drawn from a symmetric distribution. The simplex is convex, but with some probability it can be unbounded from below and then the optimization problem is ill defined. Increasing T means that the probability of this event decreases, because there are more planes and thus it is more likely that for large values of the w_i the max over t has a positive slope in the i-th direction. The exact law for this probability can be obtained by induction on N and T [6] and, as we said, it jumps in the thermodynamic limit from 1 to 0 at $N/T = 0.5$. The example of the max-loss risk measure is also helpful because it allows to stress two aspects of the problem: 1) even for finite N and T there is a finite chance that the risk measure is unbounded from below in some samples, and 2) the phase transition occurs in the thermodynamic limit when N/T is strictly smaller than 1, i.e. much before the covariance matrix develops zero modes. The very nature of the problem is that the risk measure there is simply not bounded from below. As for the ES risk measure, the threshold value $N/T = 0.5$ can be thought of as a good approximation of the actual value for many cases of practical interest (i.e. $\beta \gtrsim 0.9$), since the corrections to this limit case are exponentially small (eq. (25)).

For finite values of β we have solved numerically eqs. (21), (22) and (24) using the following procedure. We first solve the two equations (21) and (22), which always admit a solution for (v, q_0). We then plot the l.h.s. of eq. (24) as a function of $1/t$ for a fixed value of β. This function is positive at small $1/t$ and becomes negative beyond a threshold $1/t^*$. By keeping track of $1/t^*$ (numerically obtaining via linear interpolations) for each value of β we build up the phase diagram (Fig. 2, left). This diagram is in agreement with the numerical results obtained in ref. 6. We show in the right panel of Fig. 2 the divergence of the order parameter Δ versus $1/t - 1/t^*$. The critical exponent is found to be 1/2:

$$\Delta \sim \left(\frac{1}{t} - \frac{1}{t^*(\beta)}\right)^{-1/2} \ , \tag{27}$$

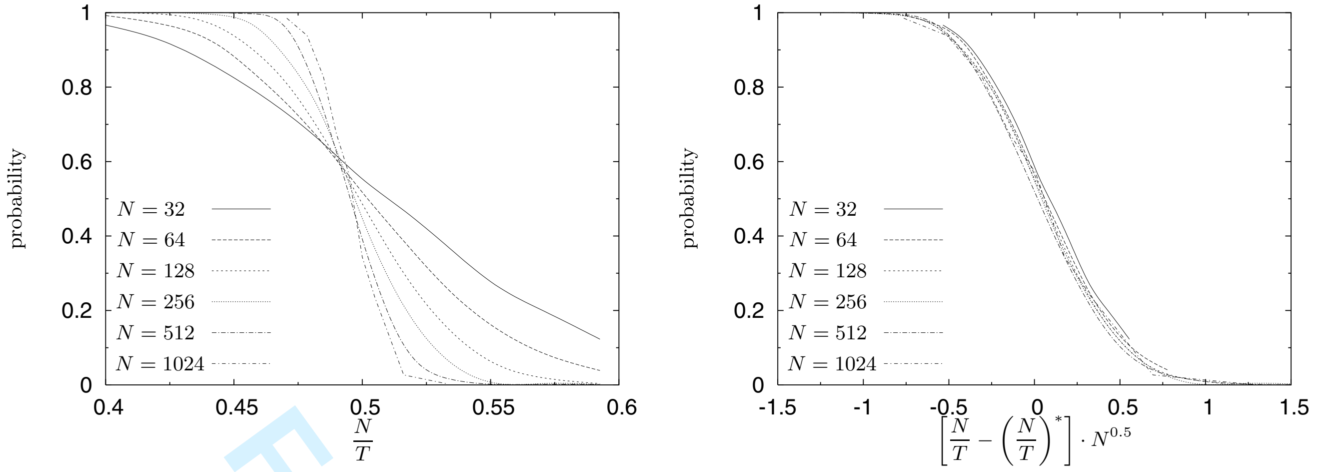

FIG. 3: Left: The probability of finding a finite solution as obtained from linear programming at increasing values of N and with $\beta = 0.8$. Right: Scaling plot of the same data. The critical value is set equal to the analytical one, $N/T = 0.4945$ and the critical exponent is $1/2$, i.e. the one obtained by Kondor et al. [6] for the limit case $\beta \to 1$. The data do not collapse perfectly, and better results can be obtained by slightly changing either the critical value or the exponent.

again in agreement with the scaling found in ref. 6. We have performed extensive numerical simulations in order to check the validity of our analytical findings. For a given realization of the time series, we solve the optimization problem (8) by standard linear programming [10]. We impose a large negative cutoff for the w's, that is $w_i > -W$, and we say that a feasible solution exists if it stays finite for $W \to \infty$. We then repeat the procedure for a certain number of samples, and then average our final results (optimal cost, optimal v, and the variance of the w's in the optimal portfolio) over those of them which produced a finite solution. In Fig. 3 we show how the probability of finding a finite solution depends on the size of the problem. Here, the probability is simply defined in terms of the frequency. We see that the convergence towards the expected 1-0 law is fairly slow, and a finite size scaling analysis is shown in the right panel. Without loss of generality, we can summarize the finite-N numerical results by writing the probability of finding a finite solution as

$$p(N, T, \beta) = f\left[\left(\frac{1}{t} - \frac{1}{t^*(\beta)}\right) \cdot N^{\alpha(\beta)}\right] , \tag{28}$$

where $f(x) \to 1$ if $x \gg 1$ and $f(x) \to 0$ if $x \ll 1$, and where $\alpha(1) = 1/2$. It is interesting to note that these results do not depend on some initial conditions of the algorithm we used to solve the problem: for a given sample, the algorithm finds in a linear time the minimum of the polytope by looking at all its vertexes exhaustively. The statistics is taken by repeating such a deterministic procedure on a large number of samples chosen at random.

In Fig. 4 (left panel) we plot, for a given value of β, the optimal cost found numerically for several values of the size N compared to the analytical prediction at infinite N. One can show that the cost vanishes as $\Delta^{-1} \sim (1/t - 1/t^*)^{1/2}$. The right panel of the same figure shows the behavior of the value of v which leads to the optimal cost versus N/T, for the same fixed value of β. Also in this case, the analytical ($N \to \infty$ limit) is plotted for comparison. We note that this quantity was suggested [5] to be a good approximation of the VaR of the optimal portfolio: We find here that $v_{\rm opt}$ diverges at the critical threshold and becomes negative at an even smaller value of N/T.

V. CONCLUSIONS

We have shown that the problem of optimizing a portfolio under the expected shortfall measure of risk by using empirical distributions of returns is not well defined when the ratio N/T of assets to data points is larger than a certain critical value. This value depends on the threshold β of the risk measure in a continuous way and this defines a phase diagram. The lower the value of β, the larger the length of the time series needed for the portfolio optimization. The analytical approach we have discussed in this paper allows us to have a clear understanding of this phase transition. The mathematical reason for the non-feasibility of the optimization problem is that, with a certain probability $p(N, T, \beta)$, the linear constraints in (9) define a simplex which is not bounded from below, thus leading to a solution which is not finite ($\Delta q \to \infty$ in our language), in the same way as it happens in the extreme case $\beta \to 1$

For Peer Review Only

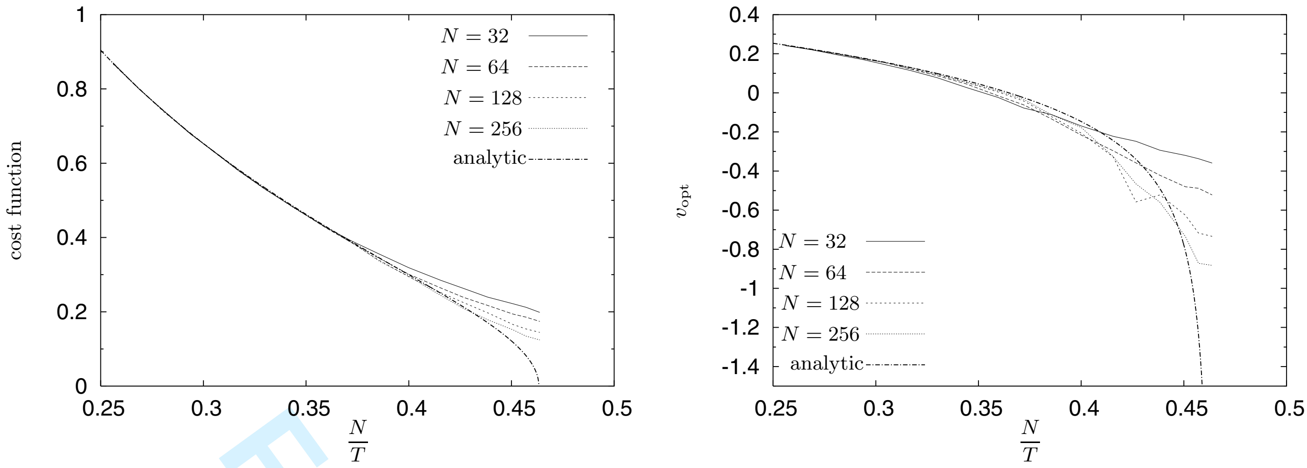

FIG. 4: Numerical results from linear programming and comparison with analytical predictions at large N. Left: The minimum cost of the optimization problem vs N/T, at increasing values of N. The thick line is the analytical solution (19). Here $\beta = 0.7$, $(N/T)^* \simeq 0.463$. Right: The optimal value of v as found numerically for several values of N is compared to the analytical solution.

discussed in [6]. From a more physical point of view, it is reasonable that the feasibility of the problem depend on the number of data points we take from the time series with respect to the number of financial instruments of our portfolio. The probabilistic character of the time series is reflected in the probability $p(N, T, \beta)$. Interestingly, this probability becomes a threshold function at large N if $N/T \equiv 1/t$ is finite, and its general form is given in (28).

These results have a practical relevance in portfolio optimization. The order parameter discussed in this paper is tightly related to the relative estimation error [6]. The fact that this order parameter has been found to diverge means that in some regions of the parameter space the estimation error blows up which makes the task of portfolio optimization completely meaningless. The divergence of estimation error is not limited to the case of expected shortfall: as shown in [6], it happens in the case of variance and absolute deviation as well, but the noise sensitivity of expected shortfall turns out to be even greater than that of these more conventional risk measures.

There is nothing surprising about the fact that if there are no sufficient data, the estimation error is large and we cannot make a good decision. What is surprising is the fact that there is a sharply defined threshold where the estimation error actually diverges.

For a given portfolio size, it is important to know that a minimum amount of data points is required in order to perform an optimization based on empirical distributions. We also note that the divergence of the parameter Δ at the phase transition, which is directly related to the fluctuations of the optimal portfolio, may play a dramatic role in practical cases. To stress this point, we can define a sort of "susceptibility" with respect to the data,

$$\chi_{ij}^t = \frac{\partial \langle w_j \rangle}{\partial x_{it}} , \tag{29}$$

and one can show that this quantity diverges at the critical point, since $\chi_{ij} \sim \Delta$. A small change (or uncertainty) in x_{it} becomes increasingly relevant as the transition is approached, and the portfolio optimization could then be very unstable even in the feasible region of the phase diagram. We stress that the susceptibility we have introduced might be considered as a measure of the effect of the noise on portfolio selection and is very reminiscent of the measure proposed in [11].

In order to present a clean, analytic picture, we have made several simplifying assumptions in this work. We have omitted the constraint on the returns, liquidity constraints, correlations between the assets, non-stationary effects, etc. Some of these can be systematically taken into account and we plan to return to these finer details in a subsequent work.

Acknowledgments. We thank O. C. Martin, and M. Potters for useful discussions, and particularly J. P. Bouchaud for a critical reading of the manuscript. S. C. is supported by EC through the network MTR 2002-00319, STIPCO, I.K. by the National Office of Research and Technology under grant No. KCKHA005.

For Peer Review Only

APPENDIX A: THE REPLICA SYMMETRIC SOLUTION

We show in this appendix how the minimum cost function corresponding to the replica-symmetric ansatz is obtained.

The 'TrLog$Q$' term in (16) is computed by realizing that the eigenvalues of such a symmetric matrix are $(q_1 + (n-1)q_0)$ (with multiplicity 1) and $(q_1 - q_0)$ with multiplicity $n-1$. Then,

$$\mathrm{Tr}\log Q = \log\det Q = \log(q_1 + (n-1)q_0) + (n-1)\log(q_1 - q_0) = n\left(\log\Delta q + \frac{q_1}{\Delta q}\right) + \mathcal{O}(n^2)\ , \tag{A1}$$

where $\Delta q \equiv q_1 - q_0$. The effective partition function in (17) depends on Q^{-1}, whose elements are:

$$(Q^{-1})^{ab} = \begin{cases} (\Delta q - q_0)/(\Delta q)^2 + \mathcal{O}(n) & \text{if } a = b \\ -q_0/(\Delta q)^2 + \mathcal{O}(n) & \text{if } a \neq b \end{cases} \tag{A2}$$

By introducing a Gaussian measure $dP_{q_0}(s) \equiv \frac{ds}{\sqrt{2\pi q_0}} e^{-s^2/2q_0}$, one can show that

$$\begin{aligned} \frac{1}{n}\log\hat{Z}(v,q_1,q_0) &= \frac{1}{n}\log\left\{\int\prod_a dx_a e^{-\frac{1}{2\Delta q}\sum_a (x^a)^2 + \gamma\sum_a (x^a+v)\theta(-x^a-v)}\int dP_{q_0}(s) e^{\frac{s}{\Delta q}\sum_a x^a}\right\} \\ &= \frac{q_0}{2\Delta q} + \int dP_{q_0}(s)\log B_\gamma(s,v,\Delta q) + \mathcal{O}(n) \end{aligned} \tag{A3}$$

where we have defined

$$B_\gamma(s,v,\Delta q) \equiv \int dx \exp\left(-\frac{(x-s)^2}{2\Delta q} + \gamma(x+v)\theta(-x-v)\right)\ . \tag{A4}$$

The exponential in (16) now reads $\exp Nn[S(q_0,\Delta q,\hat{q}_0,\Delta\hat{q}) + \mathcal{O}(n)]$, where

$$\begin{aligned} S(q_0,\Delta q,\hat{q}_0,\Delta\hat{q}) = q_0\Delta\hat{q} + \hat{q}_0\Delta q + \Delta q\Delta\hat{q} - \Delta\hat{q} - \gamma t(1-\beta)v - t\log\gamma + t\int dP_{q_0}(s)\log B_\gamma(s,v,\Delta q) \\ -\frac{t}{2}\log\Delta q - \frac{1}{2}\left(\log\Delta\hat{q} + \frac{\hat{q}_0}{\Delta\hat{q}}\right) - \frac{\log 2}{2}\ . \end{aligned} \tag{A5}$$

The saddle point equations for $\hat{q}_0$ and $\Delta\hat{q}$ allow then to simplify this expression. The free energy $(-\gamma)f_\gamma = \lim_{n\to 0}\partial\overline{Z_\gamma^n}/\partial n$ is given by

$$-\gamma f_\gamma(v,q_0,\Delta q) = \frac{1}{2} - t\log\gamma + \frac{1-t}{2}\log\Delta q + \frac{q_0 - 1}{2\Delta q} - \gamma t(1-\beta)v + t\int dP_{q_0}(s)\log B_\gamma(s,v,\Delta q)\ , \tag{A6}$$

where the actual values of v, q_0 and Δq are fixed by the saddle point equations

$$\frac{\partial f_\gamma}{\partial v} = \frac{\partial f_\gamma}{\partial q_0} = \frac{\partial f_\gamma}{\partial\Delta q} = 0\ . \tag{A7}$$

A close inspection of these saddle point equations allows one to perform the low temperature $\gamma\to\infty$ limit by assuming that $\Delta q = \Delta/\gamma$ while v and q_0 do not depend on the temperature. In this limit one can show that

$$\lim_{\gamma\to\infty}\frac{1}{\gamma}\log B_\gamma(s,v,\Delta/\gamma) = \begin{cases} s + v + \Delta/2 & s < -v - \Delta \\ -(v+s)^2/2\Delta & -v-\Delta \leq s < -v \\ 0 & s \geq -v \end{cases} \tag{A8}$$

If we plug this expression into eq. (A6) and perform the large-γ limit we get the minimum cost:

$$E = \lim_{\gamma\to\infty} f_\gamma = -\frac{q_0 - 1}{2\Delta} + t(1-\beta)v - t\int_{-\infty}^{-\Delta}\frac{dx}{\sqrt{2\pi q_0}}e^{-\frac{(x-v)^2}{2q_0}}\left(x + \frac{\Delta}{2}\right) + \frac{t}{2\Delta}\int_{-\Delta}^{0}\frac{dx}{\sqrt{2\pi q_0}}e^{-\frac{(x-v)^2}{2q_0}}x^2\ . \tag{A9}$$

We rescale $x \to x\Delta$, $v \to v\Delta$, and $q_0 \to q_0\Delta^2$, and after some algebra we obtain eq. (19).

---

[1] see P. Artzner, F. Delbaen, J. M. Eber, and D. Heath, Mathematical Finance **9,** 203–228 (1999), for an axiomatic definition of coherence.

For Peer Review Only

[2] C. Acerbi, and D. Tasche, Journal of Banking and Finance **26,** 1487–1503 (2002).
[3] R. Frey, and A. J. McNeil, (2002).
[4] G. C. Pflug, in "Probabilistic Constrained Optimization: Methodology and Applications", S. Uryasev (ed.), Kluwer Academic Publisher (2000).
[5] R. Rockafellar, and S. Uryasev, The Journal of Risk **2,** 21–41 (2000).
[6] I. Kondor, Noise sensitivity of portfolio selection under various risk measures, Lectures given at the Summer School on "Risk Measurement and Control", Rome, June 9-17, 2005; I. Kondor, S. Pafka, and G. Nagy, Noise sensitivity of portfolio selection under various risk measures, submitted to J. of Banking and Finance.
[7] see e.g. F. Guerra, F.L. Toninelli, J. Stat. Phys. **115,** 531 (2004) and references therein.
[8] M. Mézard, G. Parisi, and M. .A. .Virasoro, "Spin Glass theroy and Beyond", World Scientific Lecture Notes in Physics Vol. 9, Singapore (1987).
[9] M. Talagrand, *Spin Glasses: a Challenge for Mathematicians*, Spinger-Verlag (2002).
[10] W. H. Press, S. H. Teukolsky, W. T. Wetterling, and B. P. Flannery, "Numerical Recipes in C", Cambridge University Press (Cambridge, UK, 1992).
[11] S. Pafka, and I. Kondor, Europ. Phys. J. **B27,** 277 (2002).
[12] see [2] for the subtleties related to a discrete distribution

For Peer Review Only